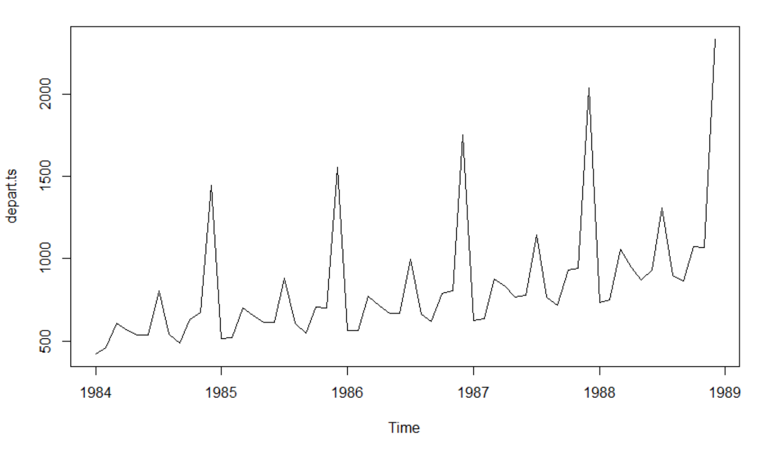
예제 1

- 자료: 예7-1. 1984년 1월부터 1988년 12월까지의 백화점 매출액
- 계절형 ARIMA 모형을 적합하고 12 선행시차에 대하여 예측하라.

- 시계열 그림으로 정상성 확인

```
> depart <- scan("D:/Data/depart.txt")
> depart.ts <- ts(depart, start = c(1984), frequency = 12)
> plot(depart.ts)
```

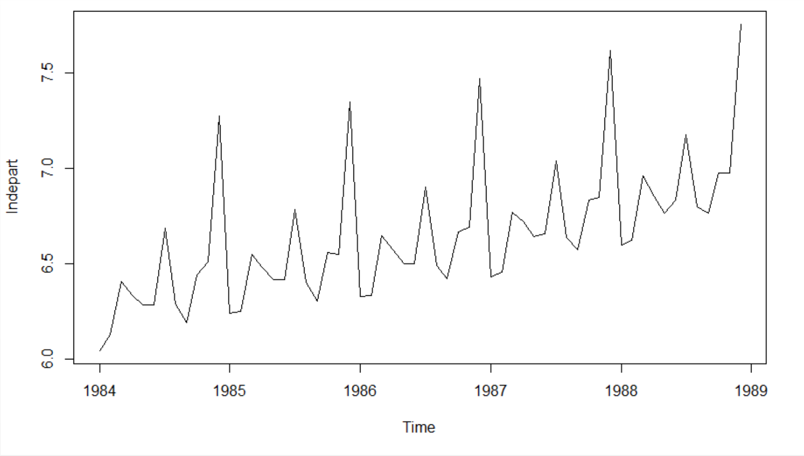


- 분산 증가
- 로그 변환 필요

- 로그 변환

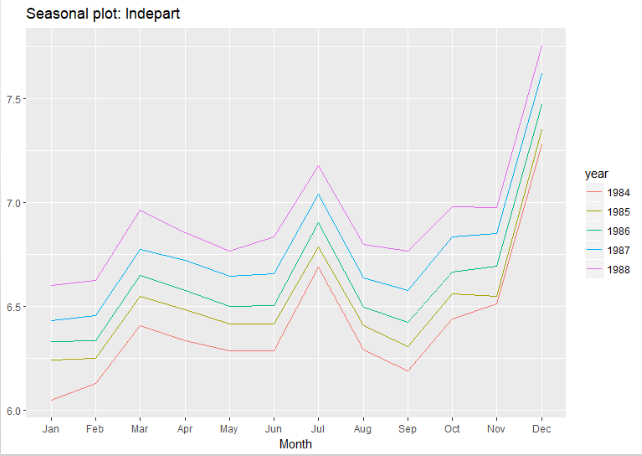
```
> lndepart <- log(depart.ts)
> plot(lndepart)
```

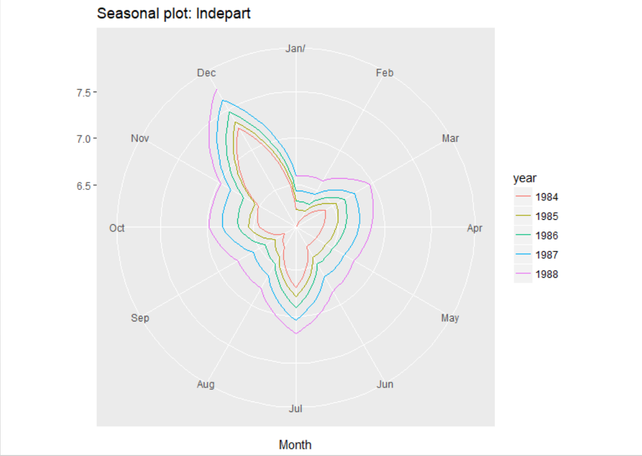
분산 안정화



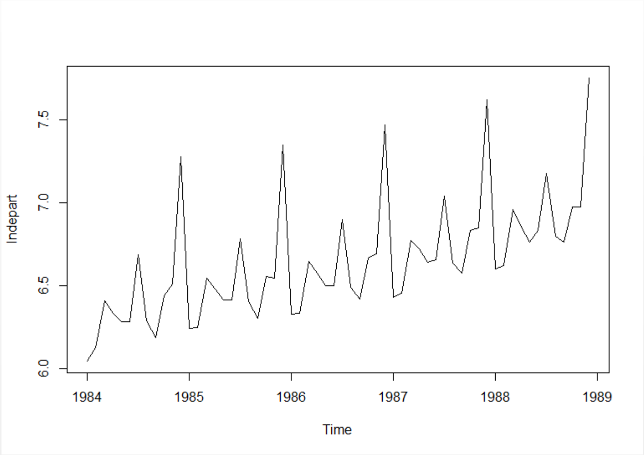
- seasonal plot

```
> library(forecast)
> ggseasonplot(lndepart)
> ggseasonplot(lndepart,polar=TRUE)
```





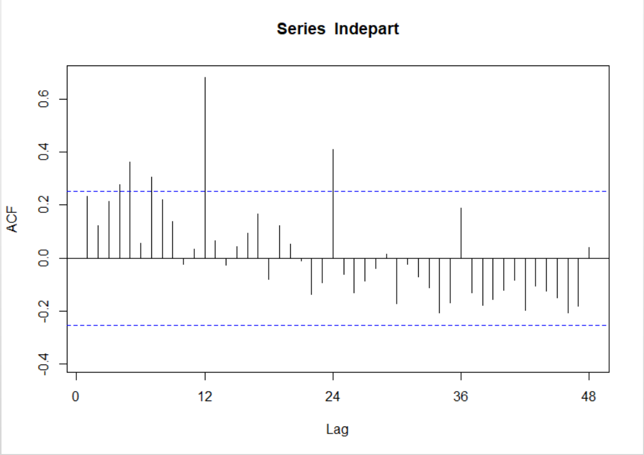
- 차분 차수 확인



시계열 그림
 - 차분이 필요한 상황

ACF
 - 차분의 필요성이 명확하지 않은 상황

```
> Acf(lndepart,lag.max=48)
```

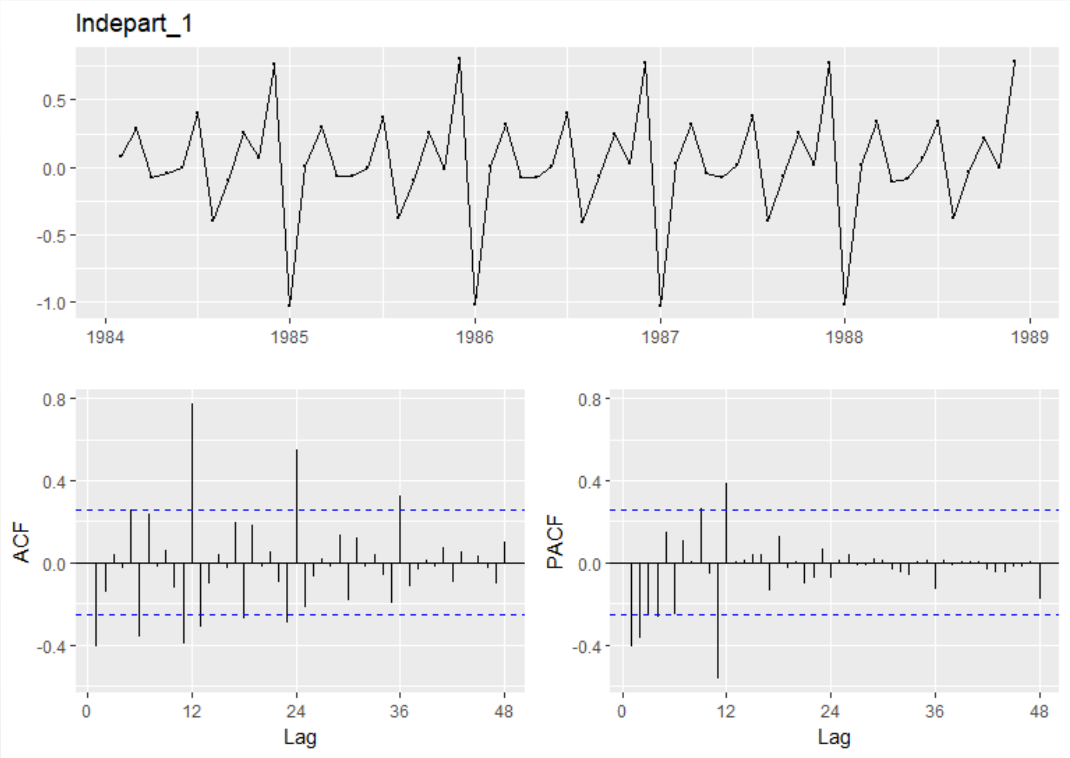



```
> library(forecast)
> ndiffs(lndepart)
[1] 1
> nsdiffs(lndepart)
[1] 0
```

차분을 시도하고 그 결과를 확인할 필요가 있음

- d=1의 경우

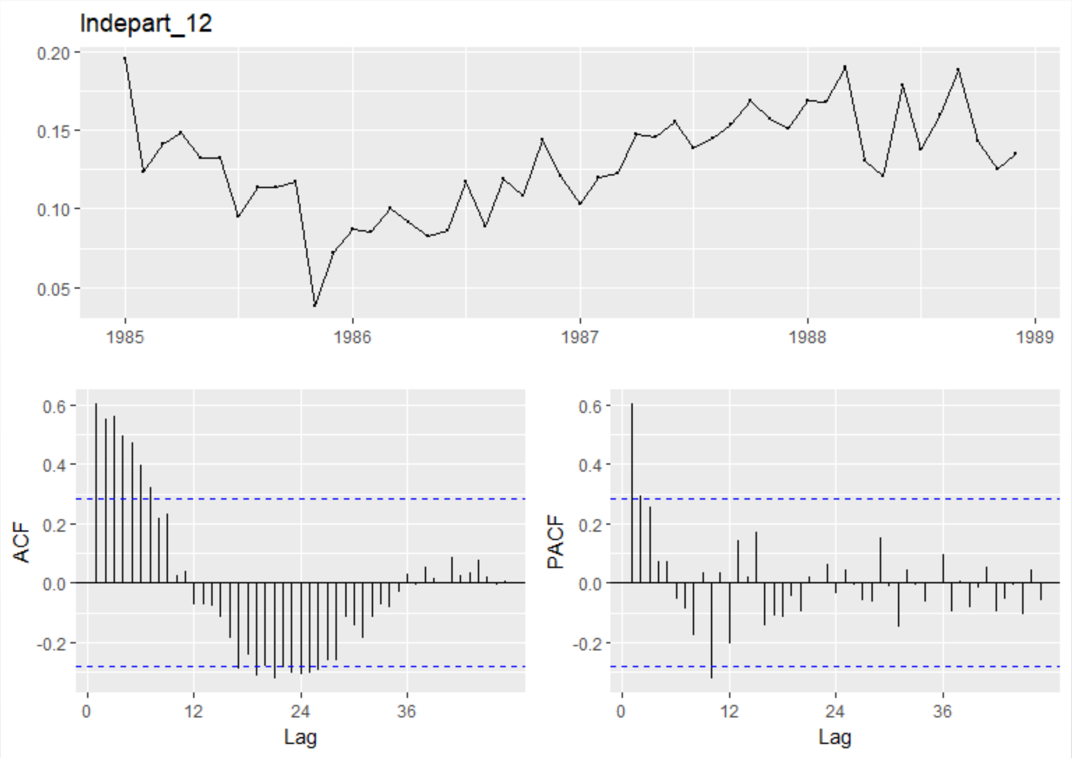
```
> lndepart_1 <- diff(lndepart)
> ggtsdisplay(lndepart_1,lag.max=48)
```



추가적인 계절 차분이 필요한 것으로 보임

- D=1의 경우

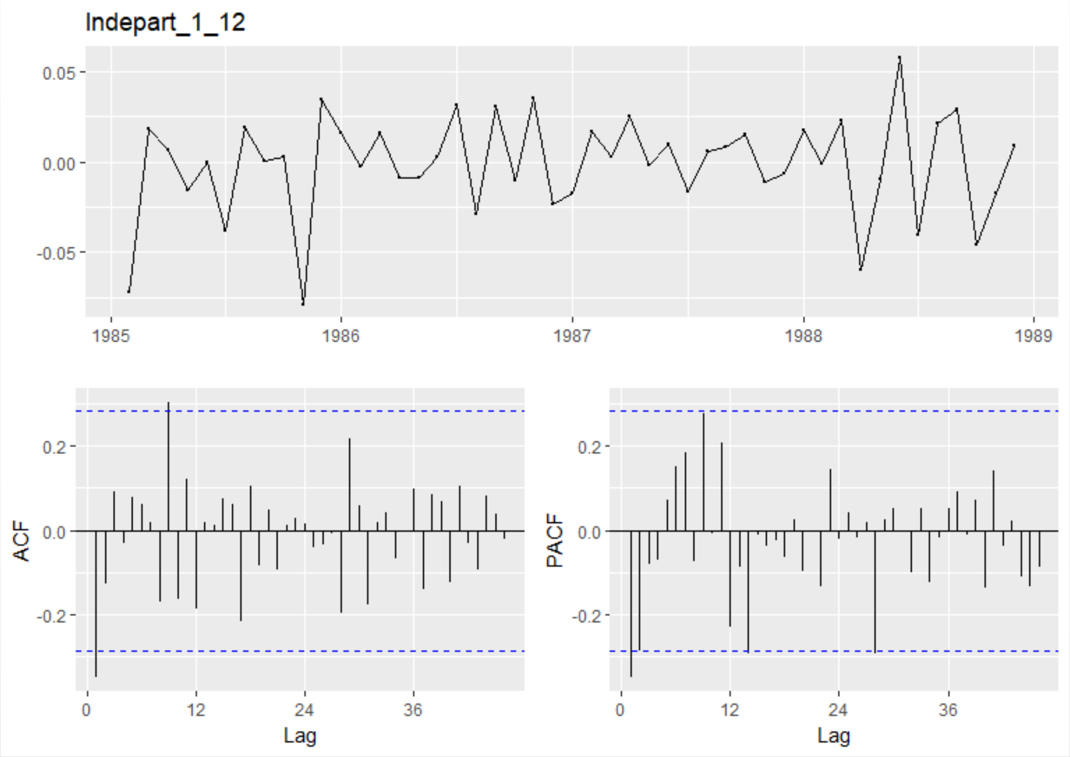
```
> lndepart_12 <- diff(lndepart,lag=12)
> ggtsdisplay(lndepart_12,lag.max=48)
```



- 추세는 계절 차분만으로도 해결되는 경우가 있음
- 이 경우에는 추가적인 일반 차분이 필요하다고 보임

- D=1, d=1의 경우

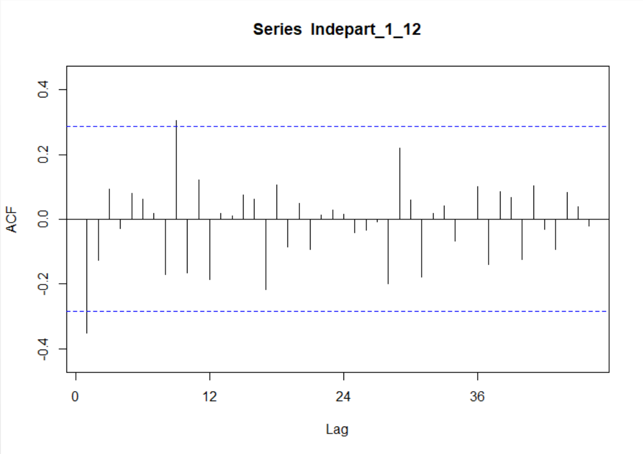
```
> lndepart_1_12 <- diff(lndepart_1,lag=12)
> ggtsdisplay(lndepart_1_12,lag.max=48)
```

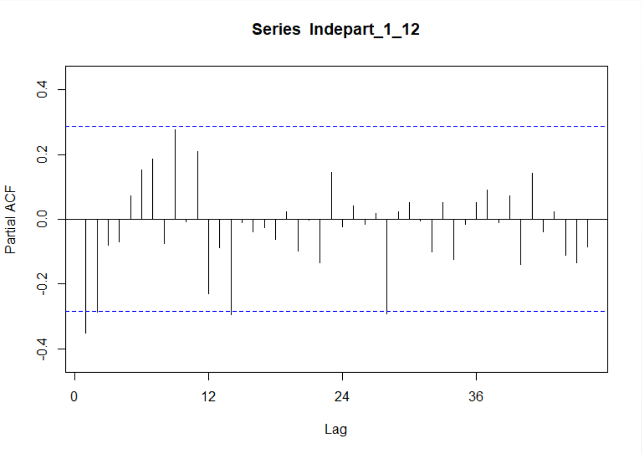


- 정상성 확보
- 최적 차분: d=1, D=1

- 모형 인식

```
> Acf(lndepart_1_12,lag.max=48)
> Pacf(lndepart_1_12,lag.max=48)
```





비계절 요소:
ACF 1시차 절단, PACF 감소 → p=0, q=1
ACF 감소, PACF 2시차 절단 → p=2, q=0

계절 요소:
12, 24, 36, 48 시차에서 모두 비유의적 → P=0, Q=0

1) 모형: $ARIMA(0,1,1)(0,1,0)_{12}$

- 모수 추정

```
> library(forecast)
> fit1 <- Arima(lndepart, order=c(0,1,1),
                seasonal=list(order=c(0,1,0),period=12))

> fit1
Series: lndepart
ARIMA(0,1,1)(0,1,0)[12]

Coefficients:
          ma1
      -0.5633
s.e.   0.1124

sigma^2 estimated as 0.0006043:  log likelihood=108.23
AIC=-212.46   AICc=-212.18   BIC=-208.76
```

- 모형 검진
  - forecast::checkresiduals( )에 의한 검진
  - 함수 tsdiag( ) 보다 더 편리한 방법

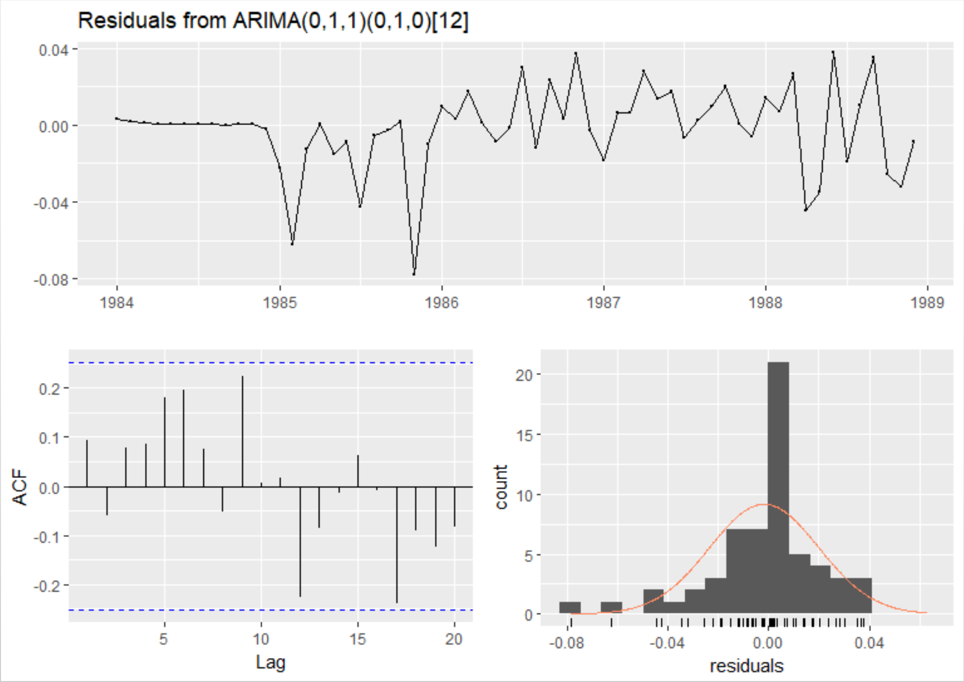
```
> checkresiduals(fit1)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,1,0)[12]
Q* = 26.262, df = 23, p-value = 0.2887

Model df: 1.   Total lags used: 24
```

- 패키지 forecast 작성자 Hyndman이 추천하는 Ljung-Box 검정에서의 K 값
  1) 비계절형 모형: K=10
  2) 계절형 모형: K=2s
     s: 주기



- 과대적합: ARIMA$(0,1,1)(0,1,0)_{12}$

```
> confint(Arima(lndepart,order=c(1,1,1),
                 seasonal=list(order=c(0,1,0),period=12)))
         2.5 %     97.5 %
ar1 -0.4675603  0.3998847
ma1 -0.8663890 -0.2253052
> confint(Arima(lndepart,order=c(0,1,2),
                 seasonal=list(order=c(0,1,0),period=12)))
         2.5 %     97.5 %
ma1 -0.9562689 -0.2298099
ma2 -0.2958022  0.3652605
```

- ARIMA$(0,1,1)(0,1,0)_{12}$ 예측 모형으로 선택 가능
- 과대적합은 비계절형 모수에만 적용

2) 모형: ARIMA$(2,1,0)(0,1,0)_{12}$

- 모수 추정

```
> fit2 <- Arima(lndepart,order=c(2,1,0),
+                 seasonal=list(order=c(0,1,0),period=12))
> fit2
Series: lndepart
ARIMA(2,1,0)(0,1,0)[12]

Coefficients:
          ar1      ar2
      -0.5269  -0.3358
s.e.   0.1474   0.1458

sigma^2 estimated as 0.0006147:  log likelihood=108.34
AIC=-210.67   AICc=-210.11   BIC=-205.12
```

비교

ARIMA$(0,1,1)(0,1,0)_{12}$:

AIC=-212.46   BIC=-208.76

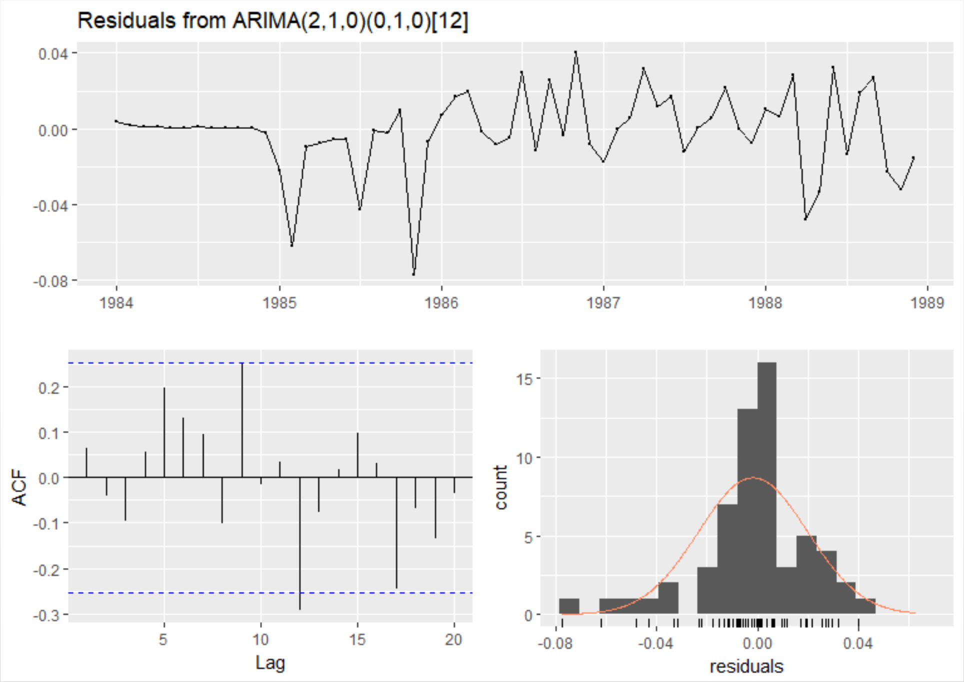
- 모형검진

```
> checkresiduals(fit2)

        Ljung-Box test

data:  Residuals from ARIMA(2,1,0)(0,1,0)[12]
Q* = 28.447, df = 22, p-value = 0.1613

Model df: 2.   Total lags used: 24
```



- 과대적합: ARIMA(2,1,0)(0,1,0)$_{12}$

```
> confint(Arima(lndepart,order=c(3,1,0),
                 seasonal=list(order=c(0,1,0),period=12)))
         2.5 %      97.5 %
ar1 -0.8739910 -0.26541395
ar2 -0.7344992 -0.07719871
ar3 -0.4476723  0.18256271
> confint(Arima(lndepart,order=c(2,1,1),
                 seasonal=list(order=c(0,1,0),period=12)))
         2.5 %    97.5 %
ar1 -0.8095983 0.3742461
ar2 -0.6154847 0.1764711
ma1 -0.9264064 0.2084493
```

- 추가된 모수 비유의적
- ARIMA(0,1,1)(0,1,0)$_{12}$ 모형 보다 큰 값의 AIC & BIC

3) auto.arima( )에 의한 모형 선택

```
> fit3 <- auto.arima(lndepart,d=1,D=1)
> fit3
Series: lndepart
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.5840  -0.4159
s.e.   0.1093   0.1946

sigma^2 estimated as 0.0005401:  log likelihood=110.29
AIC=-214.59   AICc=-214.03   BIC=-209.04
```

비교

$ARIMA(0,1,1)(0,1,0)_{12}$: AIC=-212.46 BIC=-208.76

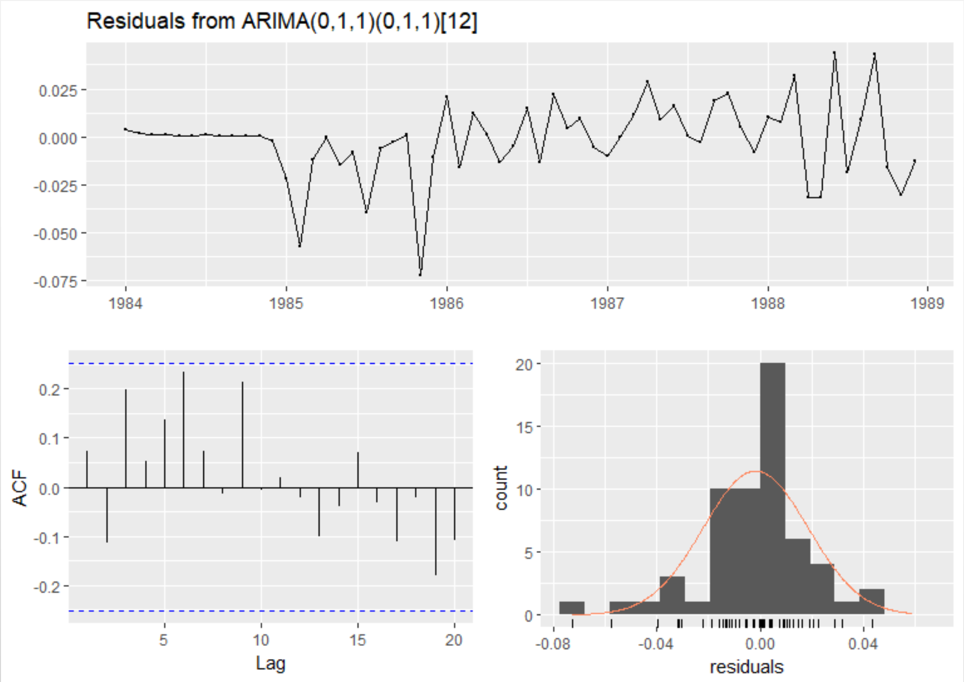
- 모형검진:

```
> checkresiduals(fit3)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,1,1)[12]
Q* = 21.548, df = 22, p-value = 0.4871

Model df: 2.   Total lags used: 24
```



- 과대적합: ARIMA(0,1,1)(0,1,1)$_{12}$

```
> confint(Arima(lndepart,order=c(1,1,1),
+              seasonal=list(order=c(0,1,1),period=12)))
          2.5 %      97.5 %
ar1  -0.4596994  0.38086082
ma1  -0.8673157 -0.26263265
sma1 -0.7964791 -0.03491842
> confint(Arima(lndepart,order=c(0,1,2),
+              seasonal=list(order=c(0,1,1),period=12)))
          2.5 %      97.5 %
ma1  -1.0150960 -0.24179869
ma2  -0.3098821  0.41114957
sma1 -0.7958091 -0.03618881
```

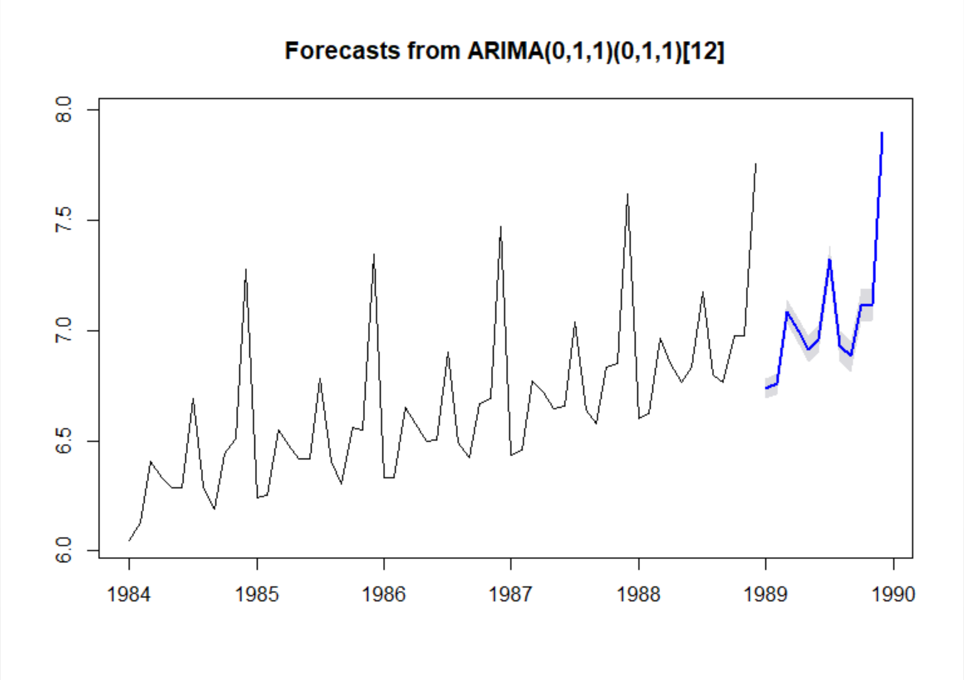
추가된 모수 비유의적

최종 모형: ARIMA(0,1,1)(0,1,1)$_{12}$

모형식: $(1-B)(1-B^{12})\log Z_t = (1-0.58B)(1-0.42B^{12})\varepsilon_t$

- 로그 변환된 자료에 대한 예측: ARIMA(0,1,1)(0,1,1)$_{12}$

```
> plot(forecast(fit3,h=12,level=95))
```



- 원 자료에 대한 예측

  - 함수 Arima( )에 lambda= 이용
  - Box-Cox 변환 모수인 λ 값을 지정하면 자료 변환 후 모형 적합
  - 로그 변환: $\lambda = 0$

```
> fit3_1 <- Arima(depart.ts,order=c(0,1,1),
             seasonal=list(order=c(0,1,1),period=12),lambda=0)
```

  - fit3_1: fit3와 동일한 결과
  - fit3_1 객체를 함수 forecast( )에 적용시키면 원 자료 크기로 예측 실시

- Box-Cox transformation

  - 양의 값을 갖는 반응변수에 대한 변환 방법
  - 주어진 자료에 가장 잘 어울리도록 반응변수를 변환
  - 반응변수 Y를 λ에 의하여 $g_\lambda(Y)$로 변환

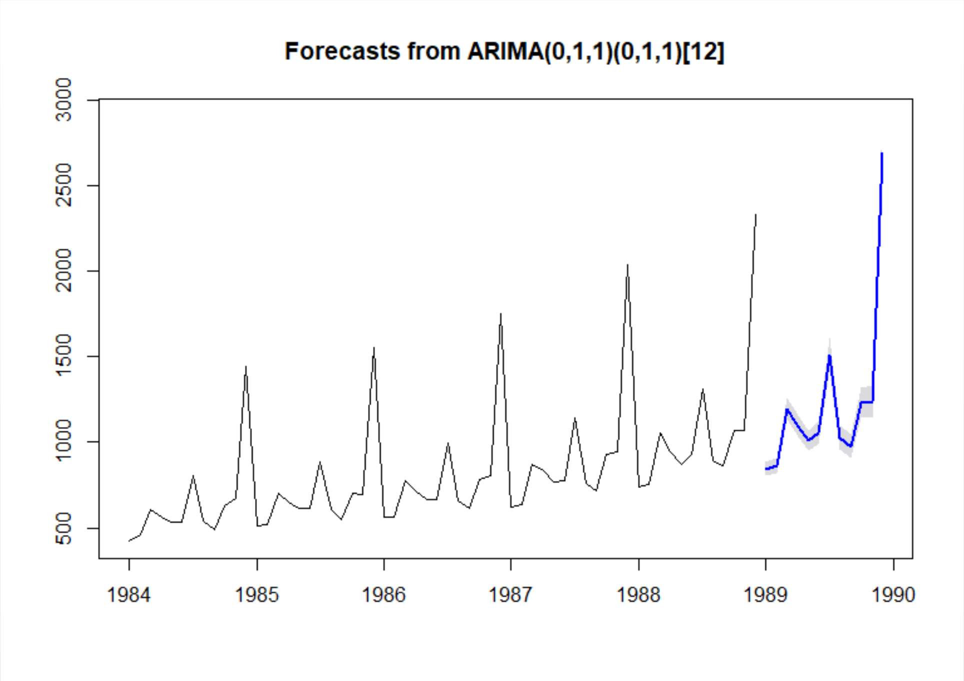
$$g_\lambda(Y) = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log Y, & \lambda = 0 \end{cases}$$

  - λ의 선택 기준은 Maximum Likelihood
  - R에서는 패키지 MASS의 함수 boxcox( ) 이용
  - 반응변수의 변환: 분석결과의 해석이 어려워짐
  - 선택된 λ → 해석 가능한 정수로

    예) $\hat{\lambda} = 0.46$ → $\sqrt{Y}$

- 원 자료에 대한 예측: ARIMA(0,1,1)(0,1,1)$_{12}$

```
> plot(forecast(fit3_1, h=12, level=95))
```



```
> summary(forecast(fit3_1,h=12,level=95))

Model Information:
Series: depart.ts
ARIMA(0,1,1)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
          ma1     sma1
      -0.5840  -0.4159
s.e.   0.1093   0.1946

sigma^2 estimated as 0.0005401:  log likelihood=110.29
AIC=-214.59   AICc=-214.03   BIC=-209.04

Error measures:
                    ME    RMSE      MAE       MPE    MAPE      MASE       ACF1
Training set -1.937472 16.9307 11.60509 -0.200563 1.36766 0.1084166 0.04358066

Forecasts:
         Point Forecast     Lo 95     Hi 95
Jan 1989       843.2496  805.6917  882.5584
Feb 1989       861.0162  819.5607  904.5686
Mar 1989      1196.5509 1134.9464 1261.4993
Apr 1989      1094.6421 1034.8703 1157.8662
May 1989      1008.7262  950.6838 1070.3124
Jun 1989      1054.5252  990.9145 1122.2193
Jul 1989      1512.3735 1417.1408 1614.0059
Aug 1989      1025.0112  957.8737 1096.8544
Sep 1989       976.3806  910.0591 1047.5353
Oct 1989      1229.0466 1142.6938 1321.9250
Nov 1989      1235.2486 1145.6798 1331.8199
Dec 1989      2690.5489 2489.6022 2907.7149
```